Descriptive Statistics in Clinical Trials

Shu-Ping Chen

2020/10/17

Descriptive Statistics in Clinical Trials

A. Postoperative pain Clinical Trial

In a double-blind, placebo-controlled study, patients with postoperative pain were randomly assigned to one of three groups: treatment A, treatment B, or placebo. Subjects enrolled in the study received a single dose of their assigned treatment when they reported moderate to severe postoperative pain. The primary outcome variable was the quantity of rescue medication used; this was recorded at hourly intervals for 24 hours after dosing. The study protocol also specified a secondary analysis to be carried out after combining the data over prespecified time intervals (0–3 hours, 4–6 hours, 7–12 hours, 13–24 hours). In this analysis, there are four repeated measurements, each computed as the average of the hourly measurements obtained during the interval. Variable description is in Table 1, and data are in files CTpainrescueWide.csv and CTpainrescueLong.csv.

Introduction

This study is a double-blind placebo-controlled study. In order to observe the effects of different drugs on the improvement of postoperative pain, patients will be randomly assigned to the drug A group, the drug B group, and the placebo group, and the patients will take the medicine of the treatment group when the postoperative pain is met. In addition, the data of this study is that 24 hours after taking the medicine, the number of medicines taken by the patient is recorded every hour, and this data is used for the first analysis. Furthermore, the second analysis will be carried out with the average of the four time-intervals of 0-3 hours, 4-6 hours, 7-12 hours, and 13-24 hours.

According to the variable description table in Appendix 1, the data includes the subject code, treatment group, and the number of rescue medications measured from 1 to 24 hours after surgery. There is no missing value in the data. The treatment groups are evenly distributed, respectively. There are 41 people in group A, 41 people in group B, and 40 people in group C. Observing the number of rescue drugs at different times, we can find that the distribution of the first hour (y1) is about normal, and the second to the twenty-fourth hour (The distribution of y2-y24) tends to be skewed to the right.

Exploratory Data Analysis

As can be seen from the table below, in the first hour after taking the drug, the drug A group took the lowest dose, while the placebo group took the highest dose. In addition, the doses of the drug A group and the drug B group 1-3 hours after taking the drug have a downward trend, but the doses of 4-12 hours fluctuate up and down, and continue to decline at 13 hours. The doses taken in the placebo group continued to fluctuate, but there was still a downward trend. Finally, the difference in the doses of the drugs taken between the three groups in the first hour and the last hour, with the drug B group reducing the most and the drug A group the least, and the last hour dose was similar to the placebo group.

Figure 1: Number of drugs using after the surgery every hour

		category							
		A		B		C		Total	
variable	Hour	mean	se	mean	se	mean	se	mean	Se
time	1	2.80	1.76	3.10	1.89	3.13	1.81	3.01	1.82
	2	2.32	1.97	2.27	1.96	2.55	1.97	2.38	1.96
	3	2.15	1.73	1.71	1.99	2.33	1.95	2.06	1.90
	4	1.83	1.70	1.66	1.65	2.63	1.88	2.03	1.78
	5	1.93	1.68	1.73	1.61	2.35	1.97	2.00	1.76
	6	1.80	1.45	1.93	1.98	2.28	1.84	2.00	1.77
	7	1.66	1.37	1.41	1.52	2.05	2.00	1.70	1.65
	8	1.71	1.47	1.56	1.34	2.05	1.87	1.77	1.57
	9	1.61	1.66	1.39	1.26	2.03	1.80	1.67	1.60
	10	1.78	1.39	1.61	1.59	2.00	1.71	1.80	1.56
	11	1.46	1.42	1.59	1.64	2.00	1.99	1.68	1.70
	12	1.44	1.21	1.61	1.55	1.78	1.61	1.61	1.46
	13	1.61	1.50	1.59	1.47	2.18	2.04	1.79	1.69
	14	1.46	1.52	1.32	1.42	2.05	1.58	1.61	1.53
	15	1.41	1.22	1.32	1.33	1.60	1.35	1.44	1.30
	16	1.37	1.24	1.37	1.20	1.88	1.57	1.53	1.36
	17	1.44	1.52	0.90	1.09	1.43	1.50	1.25	1.39
	18	1.10	1.36	1.05	1.02	1.08	1.19	1.07	1.19
	19	0.85	1.11	1.24	1.36	1.50	1.30	1.20	1.28
	20	1.41	1.48	1.46	1.63	1.15	1.12	1.34	1.42
	21	1.37	1.28	1.59	1.52	1.63	1.66	1.52	1.48
	22	1.34	1.42	1.07	1.35	2.03	2.01	1.48	1.65
	23	1.49	1.52	1.05	1.00	1.85	2.03	1.46	1.59
	24	1.90	1.80	1.71	1.63	2.00	1.93	1.87	1.78
	change	-0.90	1.84	-1.39	2.19	-1.13	1.68	-1.14	1.91

From the figure below, you can find the difference in the average number of drugs used by patients in the A (drug A) group, B (drug B) group, and C (placebo) group in different hours. The patients in the placebo group used more drugs on average in almost every hour than the other two groups, while the average number of drugs used by the patients in the drug A group and the drug B group was similar in most hours.

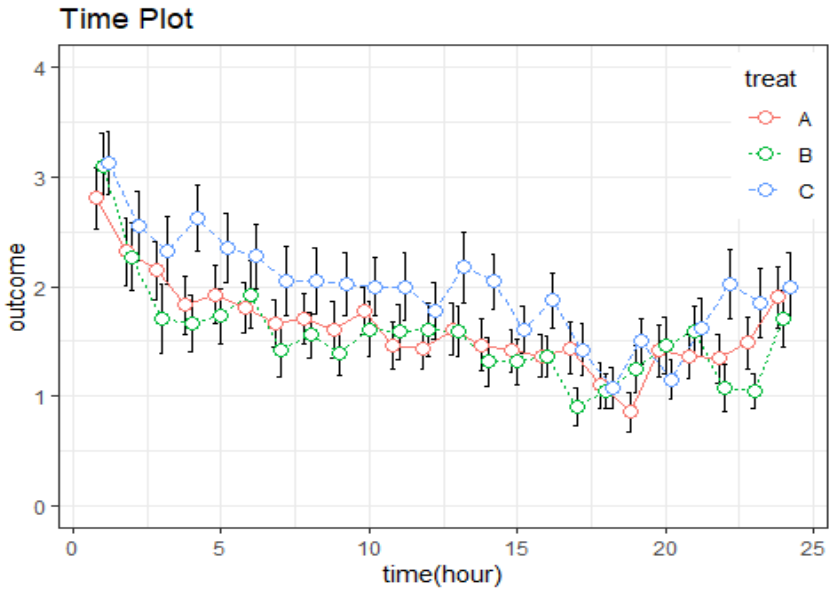


Figure 1: Change of the number of drugs taking in different treatment groups

As can be seen from the figure below, the average dose of drugs in the four periods analyzed in the second analysis, the average dose of the placebo group is higher than that of the other two groups, and the average dose of drugs between the other two groups There is not much difference in the dosage taken.

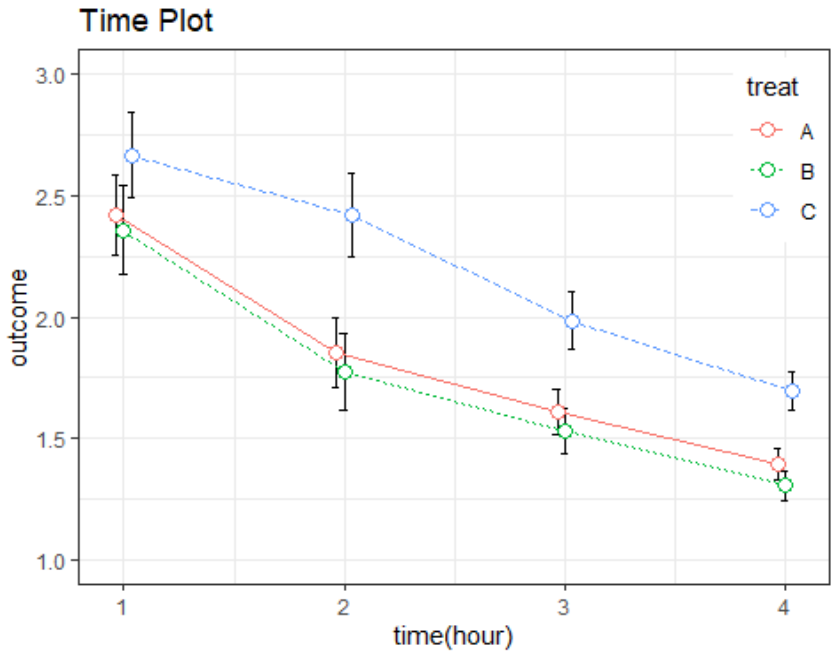


Figure 2: Change of the number of drugs taking in different treatment groups in each period

It can be seen from Figure 3 that, 1-3 hours after taking the drug, the drug doses of the patients in the drug A and drug B groups are lower than the placebo group; 4-6 hours after taking the drug, the drug A and the dose of the drug taken by the patients in the drug group B increased, while the dose taken by the placebo group decreased. The three treatment groups took the same dose at this time; 7-12 hours after taking the drug, the doses of the drugs taken by the patients in the A and drug B groups generally decreased, while the doses in the placebo group increased; 13-24 hours after taking the drugs, for patients in the three treatment groups, the dose of medications taken continued to decrease, with the placebo group having the largest decrease. It can also be seen from the figure that 4-6 hours and 13-24 hours after taking the drug, there are outliers in the drug A and drug B groups, and the distribution of the different doses of the two groups is relatively concentrated; and the placebo group had outliers 4-24 hours after taking the medicine, and its full range was larger than the other two groups.

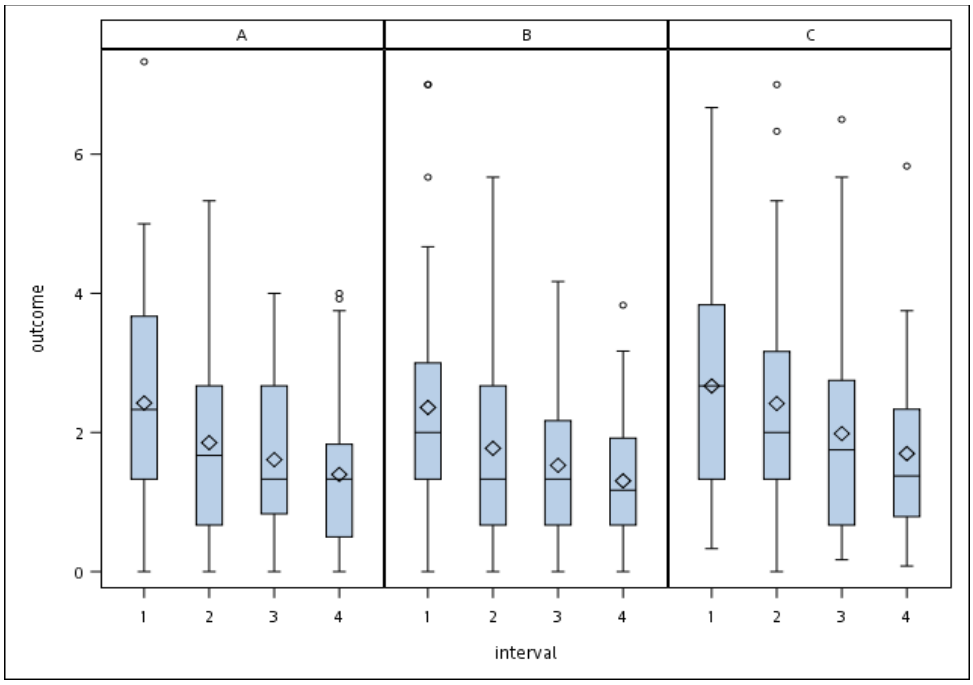


Figure 3: The dosage of medicament in 4 periods of different test groups Boxplot

From Figure 4, we can see the difference between the average dose of the test group in the first hour and the 24th hour in each group. On the whole, the average dosage of medicines in the three groups showed a downward trend. Among them, the medicine A group had the smallest decline, and the medicine B group had the largest decline.

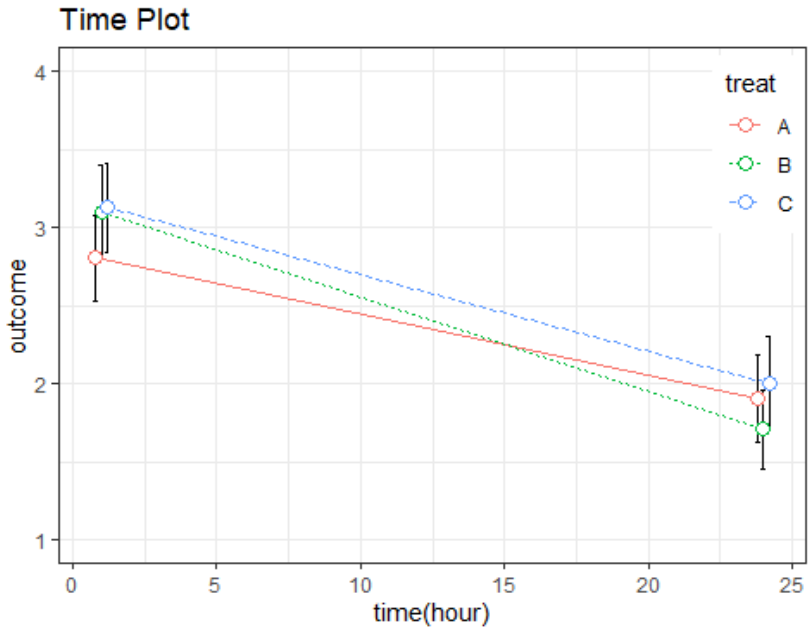


Figure 4: Drug dosage before and after 24 hours in different test groups Plot

Method

MANOVA

The test is divided into three drug groups, and the dosage of the drugs is regarded as a continuous form. Therefore, to determine whether there is a difference between the test groups, the MANOVA test is used, and the follow-up analysis is performed on the basis of significance level as 0.05.

Hypothesis Testing:

$$\begin{cases} H_0: \mu_1 = \mu_2 = \ldots = \mu_k \\ H_1: not\ H_0 \end{cases}$$

Through the MANOVA test, you can observe whether the average values of different variables are significantly different. Among them, the response variable is the number of medications taken every hour from the 1st hour to the 24th hour, and the explanatory variable is the treatment group. As can be seen from the table below, there is no significant difference in the mean of almost all explanatory variables, and only the mean of the number of drugs taken at the 4th hour and the 22nd hour is significantly different.

Table 2: MANOVA for every hour after surgery

Dependent	HypothesisType	Source	DF	SS	MS	FValue	ProbF
y1	3	treat	2	2.568	1.284	0.385	0.681
y2	3	treat	2	1.829	0.914	0.236	0.790
y3	3	treat	2	8.214	4.107	1.146	0.321
y4	3	treat	2	21.469	10.735	3.525	0.033
y5	3	treat	2	8.071	4.035	1.305	0.275
y6	3	treat	2	4.805	2.403	0.766	0.467
y7	3	treat	2	8.306	4.153	1.530	0.221
y8	3	treat	2	5.088	2.544	1.028	0.361
y9	3	treat	2	8.398	4.199	1.663	0.194
y10	3	treat	2	3.097	1.548	0.629	0.535
y11	3	treat	2	6.386	3.193	1.111	0.333
y12	3	treat	2	2.286	1.143	0.534	0.588
y13	3	treat	2	8.977	4.488	1.583	0.210
y14	3	treat	2	12.142	6.071	2.666	0.074
y15	3	treat	2	1.669	0.835	0.491	0.613
y16	3	treat	2	6.969	3.485	1.925	0.150
y17	3	treat	2	7.641	3.820	1.998	0.140
y18	3	treat	2	0.049	0.024	0.017	0.983
y19	3	treat	2	8.596	4.298	2.711	0.071
y20	3	treat	2	2.295	1.147	0.561	0.572
y21	3	treat	2	1.588	0.794	0.357	0.701
y22	3	treat	2	19.451	9.726	3.722	0.027

y23	3	treat	2	13.049	6.524	2.648	0.075
y24	3	treat	2	1.804	0.902	0.281	0.756
change	3	treat	2	4.890	2.445	0.665	0.516

Because the drug doses taken at different times (y1-y24) have a right-skewed trend as a whole, log transformation is performed on the data, and after conversion, the MANOVA test is performed again, and the alpha is set to 0.05. From the table below, we can see that only the 13th, 14th, and 16th hour averages are significantly different between the groups, and there are no significant differences for the rest.

Table 3: MANOVA after log transformation for every hour after surgery

Dependent	DF	SS	FValue	ProbF
logy1	2	0.217	0.297	0.744
logy2	2	0.235	0.271	0.763
logy3	2	1.374	1.572	0.212
logy4	2	2.168	2.939	0.057
logy5	2	0.615	0.779	0.461
logy6	2	0.505	0.643	0.528
logy7	2	0.836	1.186	0.309
logy8	2	0.499	0.721	0.488
logy9	2	0.989	1.506	0.226
logy10	2	0.395	0.569	0.568
logy11	2	0.545	0.755	0.472
logy12	2	0.320	0.524	0.594
logy13	2	0.605	0.857	0.427
logy14	2	1.937	3.134	0.047
logy15	2	0.080	0.150	0.861
logy16	2	1.656	3.027	0.052
logy17	2	0.733	1.408	0.249
logy18	2	0.019	0.047	0.954
logy19	2	0.719	1.531	0.221
logy20	2	0.453	0.814	0.445
logy21	2	0.301	0.517	0.598
logy22	2	1.780	2.729	0.069
logy23	2	1.661	2.788	0.066
logy24	2	0.131	0.159	0.854

Through the MANOVA test, you can observe whether the average values of different variables are significantly different. Among them, the number of

medications taken during the four periods of the response variable is in order: 0-3 hours, 4-6 hours, 7-12 hours, 13-24 hours, and the explanatory variable is the treatment group. As can be seen from the table below, there is no significant difference in the mean of all explanatory variables.

Table 4: MANOVA for every interval after surgery

Dependent	DF	SS	FValue	ProbF
interval1	2	2.14	0.401	0.670
interval2	2	9.94	2.209	0.114
interval3	2	4.77	1.499	0.228
interval4	2	3.38	1.605	0.205

After merging the pre-planned time intervals, perform a secondary analysis to observe in advance whether the time interval has a biased distribution trend, and observe its distribution first.

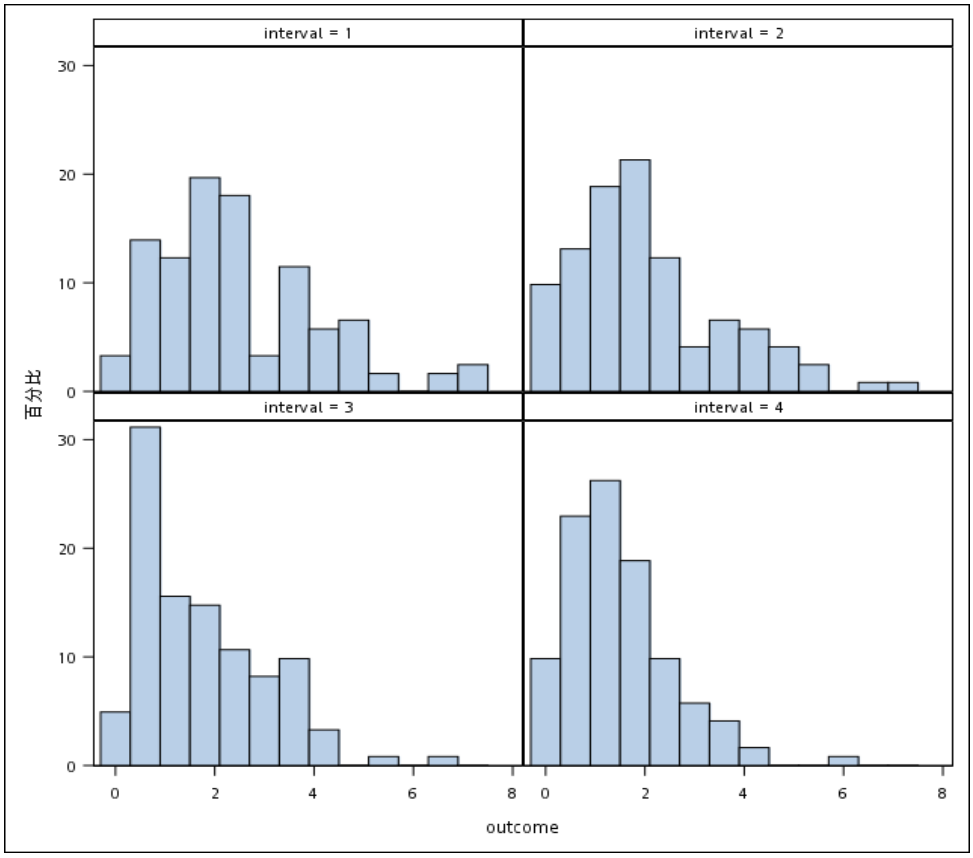


Figure 5: The dosages taking in different interval Histogram

As can be seen from the above figure, the dose distribution of drugs taken at different time intervals has a right-skewed trend, so log transformation of the data is performed. After the transformation, the MANOVA test was performed again, and

the p-values were all greater than 0.05, which can be inferred that there is no significant difference between different groups.

Table 5: MANOVA after log transformation for every interval after surgery

Dependent	DF	SS	FValue	ProbF
loginterval1	2	0.151	0.151	0.860
loginterval2	2	1.909	1.876	0.158
loginterval3	2	0.526	0.404	0.668
loginterval4	2	1.242	0.840	0.434

- Constraints: The MANOVA test cannot effectively analyze the bias, and this analysis is performed on the mean, so it is impossible to know whether there is a difference in the variance.

Conclusion

The data in this study has a serious right-skewed trend. As we directly analyze the data by MANOVA, it can be seen that only a few response variables have significant differences between the groups. In order to deal with the problem of data right-skewed, we analyze again by using MANOVA analysis after log transformation. Finally, there are still only a few means of response variables that are significantly different between groups. Therefore, it can be inferred through statistical methods that taking drugs cannot effectively reduce postoperative pain.

B. Statistical Design: CRD

Consider a Phase II trial in patients with schizophrenia. A balanced design was used in the trial with the total sample size of 240 patients. The patients were randomly allocated to four trials arms: Placebo, Low dose (40 mg), Medium dose (80 mg) and High dose (120 mg). The primary endpoint was the change from baseline in the Positive and Negative Syndrome Scale (PANSS) total score to Week 4. This endpoint was assumed following a normal distribution, and a lower value at the Week 4 visit indicated beneficial effect. The one-sided Type I error rate in this clinical trial was set to $\alpha = 0.025$. Gender and baseline PANSS were included in the data set because they were likely to influence the primary endpoint, e.g., the PANSS total score at baseline was expected to be strongly correlated with the change from baseline to Week 4. It is also worth mentioning that there were lots of missing observations in the trial. The early dropout rate was over 15% in all trial arms. For the sake of simplicity, the missing values of the primary endpoint were imputed using the basic LOCF (last observation carried forward) method. Variable description is in Table 2, and data is in file CTSchizoDosingFindingCRDBalance.csv.

Introduction

Conduct a two-stage trial on patients suffering from schizophrenia. In this trial, a balanced design was used, and the total number of samples was 240 patients. Patients will be randomly assigned to one of the following four test groups: placebo group, low-dose group (40 mg), medium-dose group (80 mg), and high-dose group (120 mg). The main test indicator is the change in the total score of the Active and Negative Symptom Scale (PANSS) from baseline to the fourth week. In this clinical trial, the single-tailed type one error was set significance level to 0.025.

Observing the variable description table in Table 2 in the appendix again, the data includes the patients' code, treatment group, gender, and the measured values at the beginning of the experiment and the fourth week after the experiment. The groups are evenly distributed, with 60 persons in each group, male and female. The ratio is also not much different. The distribution of measured values at the beginning of the experiment and the fourth week after the experiment is approximately normal. The data is processed using the Loss Value of the Basic Last Observed Value Forward Method (LOCF), so there is no missing value here.

Exploratory Data Analysis

According to the table below, the variable, change, is the amount of change before and after treatment. Firstly, we observe the gender. Under different treatment groups, males in the placebo group have an average smaller PANSS than the other groups, and there is no significant difference in PANSS between the rest of the males and females. At baseline, the average PANSS of the middle-dose group and the high-dose group was lower than that of the placebo group and the low-dose group. The PANSS in the 4th week after treatment was the opposite. Therefore, the amount of change of the average of the is larger in the middle-dose group and the high-dose group.

Table 6: PANSS value in different variables and categories

		Group									
		Placebo		Low dose		Medium dose		High dose		Total	
Variable	Category	mean	se	mean	se	mean	se	mean	se	mean	se
gender	Female	64.89	42.24	67.83	41.23	66.95	39.35	66.13	40.76	66.40	40.75
	Male	67.28	42.67	65.13	41.75	67.21	39.85	66.87	39.25	66.56	40.73
status	baseline	88.35	9.43	88.07	9.85	84.27	10.67	84.62	10.60	86.33	10.26
	change	10.60	15.10	11.52	12.46	16.33	13.37	15.13	14.30	13.40	13.97
	outcome	98.95	19.67	99.58	17.07	100.60	19.26	99.75	19.30	99.72	18.74

It can be seen from the figure below that at baseline, there was no significant difference in PANSS between different treatment groups, and the distribution of PANSS became scattered in the 4th week after treatment. Among them, the PANSS

distribution of the group taking placebo and medium dose became more scattered in the 4th week after treatment; the group taking low dose and high dose was distributed among the four treatment groups in the 4th week after treatment more concentrated.

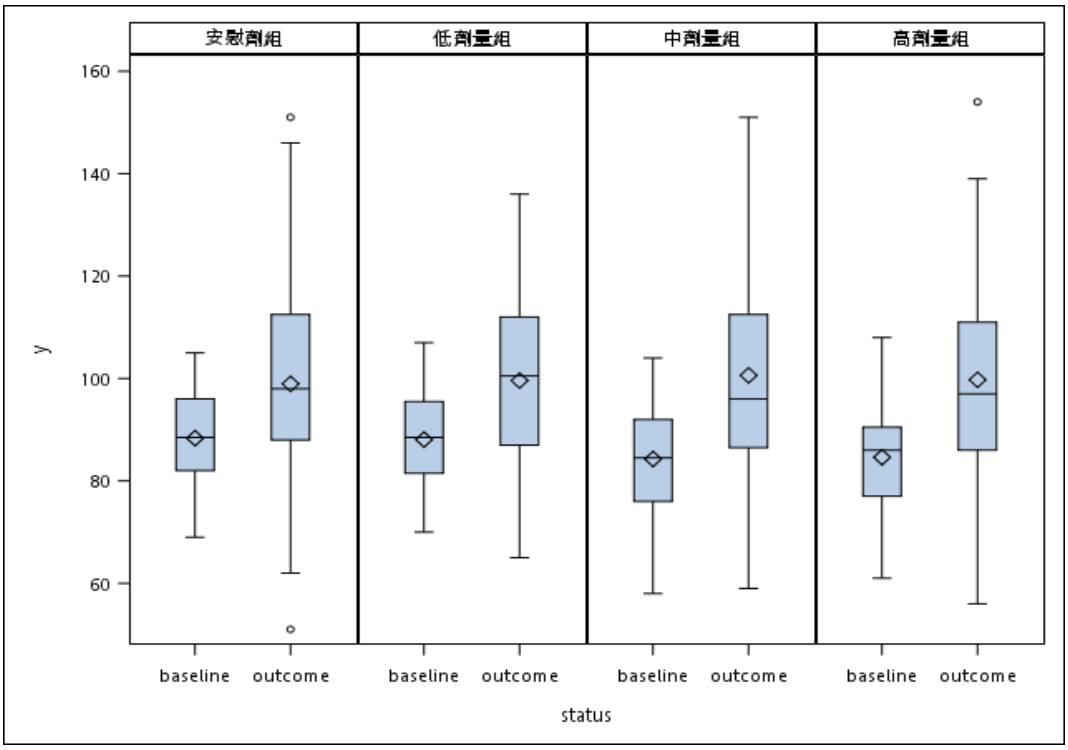


Figure 6: PANSS value for four groups in different time Boxplot

The box plot of the PANSS difference value of the different treatment groups by gender at the 4th week after treatment and the baseline can be seen from the figure below. The interquartile range of the PANSS change value of women taking placebo is more concentrated than that of men, but the total distance is higher; men's are scattered, and men have lower changes than women on average; women who take low doses have more changes in PANSS than men; women who take the middle dose group have higher changes than men on average, and there are outliers in the changes in both groups. In the high-dose group, the variation of females was more dispersed than that of males.

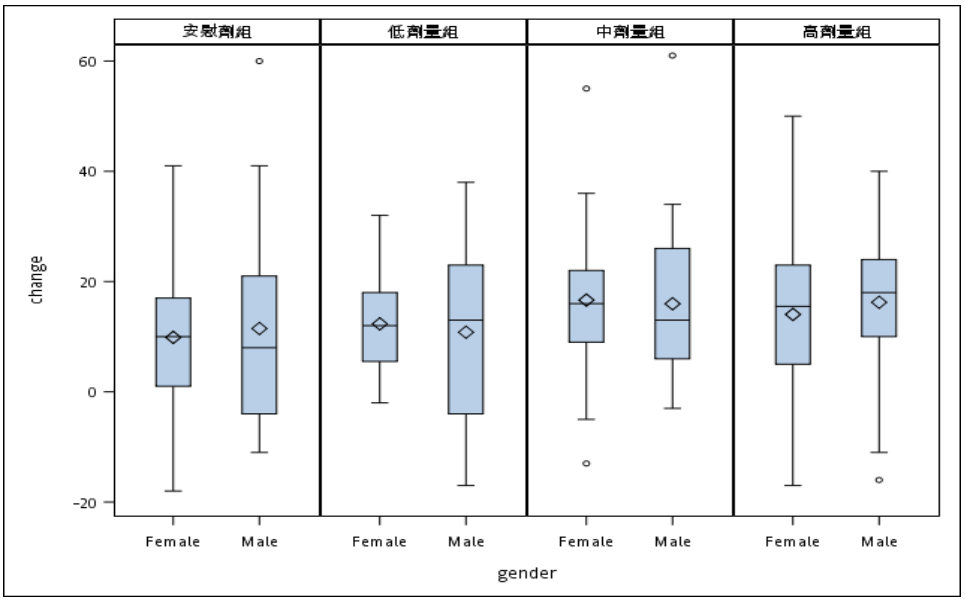


Figure 7: Difference of PANSS value for four groups in different time Boxplot

From the figure below, we can know the amount of change in PANSS from baseline (time=0) to the fourth week after the test. It can be seen from the results measured at these two time points that the difference between the four groups of test groups is not too great, but the range of changes in the test groups with doses of 80 and 120 is larger than that of the other two groups.

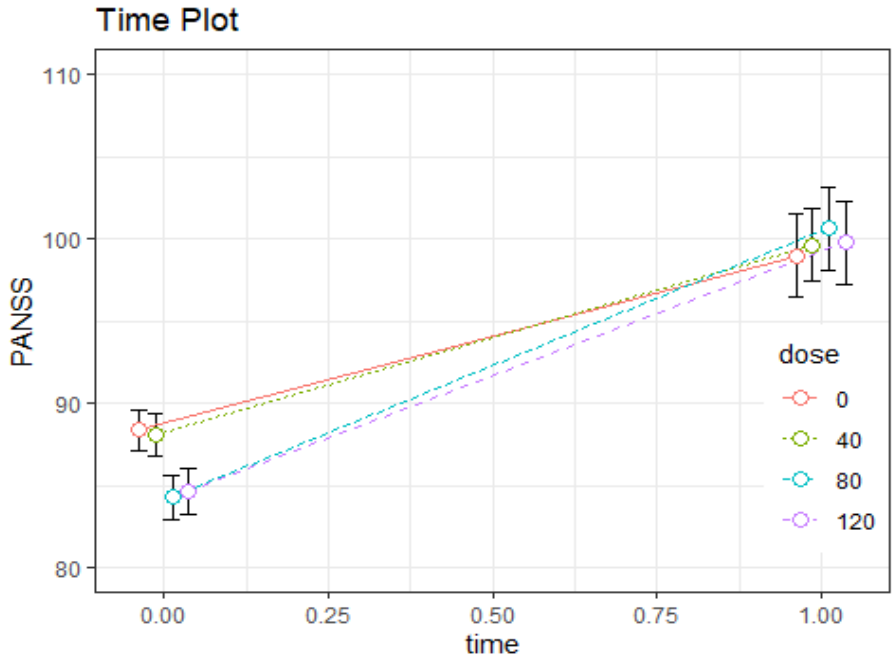


Figure 8: Change of PANSS value in different time

Method

ANOVA

Hypothesis Testing:

$$\begin{cases} H_0: the\ mean\ is\ equal\ in\ each\ group \\ H_1: not\ H_0 \end{cases}$$

It can be seen from the table below that through the ANOVA test, there is no significant difference between the treatment group, gender, and the interaction direction of the two variables. Therefore, it can be inferred that there is no significant effect on the change value of PANSS between the groups. However, because of the p-value of the difference between the treatment groups is close to 0.05, the subsequent multiple comparison method will be analyzed for the treatment group.

Table 7: ANOVA test of variable, change

Source	df	ANOVA SS	SS	F-Value	Pr > F
dose	3	1379.75	459.92	2.37	0.072
gender	1	6.73	6.73	0.03	0.853
dose*gender	3	144.75	48.25	0.25	0.863

Scheffe

It can be seen from the following table that through Scheffe method, it can be known that there is no significant difference between treatment groups, so it is concluded that there is no significant effect on the change value of PANSS between groups.

Table 8: Scheffe Multiple Comparison for variable, change

Dose Comparison	Mean of the difference	Simultaneous 95% C.I Lower Bound	Simultaneous 95% C.I Upper Bound	Significance
80 - 120	1.200	-5.97	8.37	.
80 - 40	4.817	-2.35	11.98	.
80 - 0	5.733	-1.44	12.90	.
120 - 80	-1.200	-8.37	5.97	.
120 - 40	3.617	-3.55	10.79	.
120 - 0	4.533	-2.63	11.70	.
40 - 80	-4.817	-11.98	2.35	.
40 - 120	-3.617	-10.79	3.55	.

40 - 0	0.917	-6.25	8.09	.
0 - 80	-5.733	-12.90	1.44	.
0 - 120	-4.533	-11.70	2.63	.
0 - 40	-0.917	-8.09	6.25	.

Bonferroni

As more groups are tested, the actual probability of type I error will increase. To avoid type I error, the Bonferroni method in the multiple comparison method will be used. This method divides the initially set significance level by the total executed average number to find the new significance level.

It can be seen from the table below that through the Bonferroni method, it can be known that there is no significant difference between the treatment groups, so it is inferred that there is no significant effect on the change of PANSS value between the groups.

Table 9: Bonferroni Multiple Comparison for variable, change

Dose Comparison	Mean of the difference	Simultaneous 95% C.I Lower Bound	Simultaneous 95% C.I Upper Bound	Significance
80 - 120	1.200	-5.57	7.97	.
80 - 40	4.817	-1.96	11.59	.
80 - 0	5.733	-1.04	12.51	.
120 - 80	-1.200	-7.97	5.57	.
120 - 40	3.617	-3.16	10.39	.
120 - 0	4.533	-2.24	11.31	.
40 - 80	-4.817	-11.59	1.96	.
40 - 120	-3.617	-10.39	3.16	.
40 - 0	0.917	-5.86	7.69	.
0 - 80	-5.733	-12.51	1.04	.
0 - 120	-4.533	-11.31	2.24	.
0 - 40	-0.917	-7.69	5.86	.

model

Observing the following table shows that the difference between dose and gender does not affect the change value of PANSS, so it can be inferred that the relationship is not linear. The residual figure of the model shows that there is no obvious trend in the residual error, so it can be inferred that there is no secondary relationship between the change value of PANSS, dose group and gender.

Table 10: ANOVA Table of variable, change

Source	df	Type I SS	Mean square	F value	Pr > F
dose	3	1379.75	459.92	2.37	0.072
gender	1	9.88	9.88	0.05	0.822
dose*gender	3	141.60	47.20	0.24	0.866

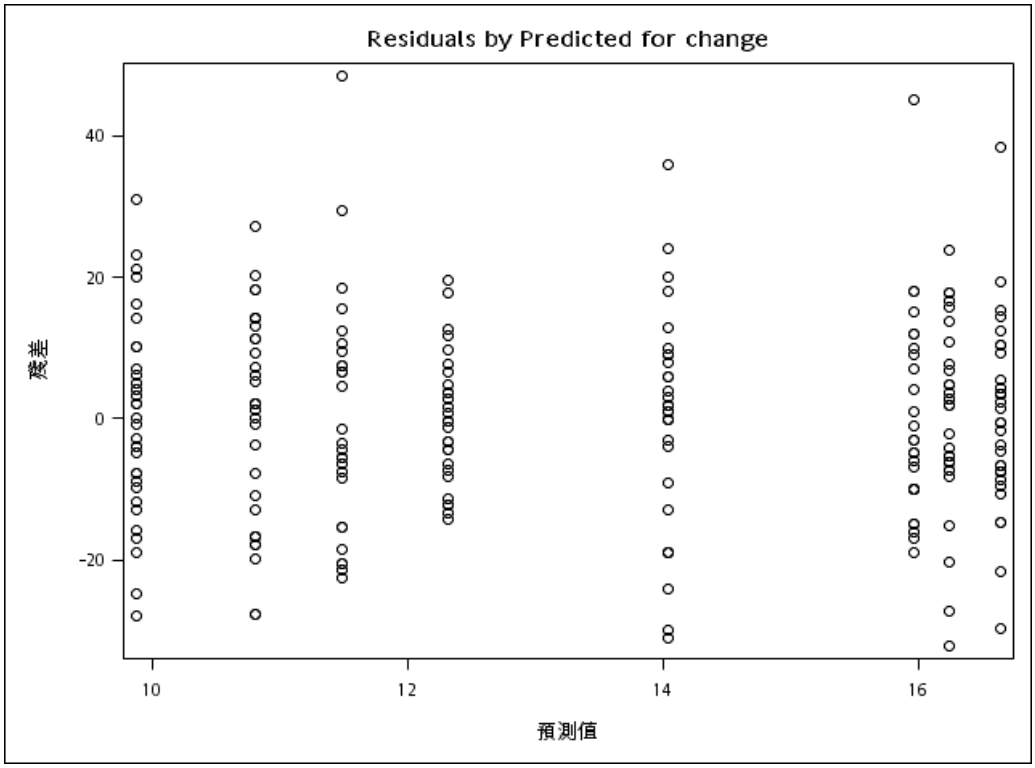


Figure 9: Model residual plot of variable, change

Observing the following table shows that there are significant differences in PANSS between baseline and treatment group in the 4th week after treatment, but there is no significant difference in gender and interaction. Observing the residual plot again, it can be found that there is no obvious trend, so it can be inferred that the first after treatment, the 4-week PANSS has a linear relationship with baseline and treatment group, rather than a quadratic relationship.

Table 11: ANOVA Table of variable, outcome

Source	df	Type I SS	Mean square	F value	Pr > F
baseline	1	38769.9	38769.9	208.22	<.0001
dose	3	2016.1	672.0	3.61	1.4E-2
gender	1	12.2	12.2	0.07	0.79
dose*gender	3	121.7	40.6	0.22	0.88

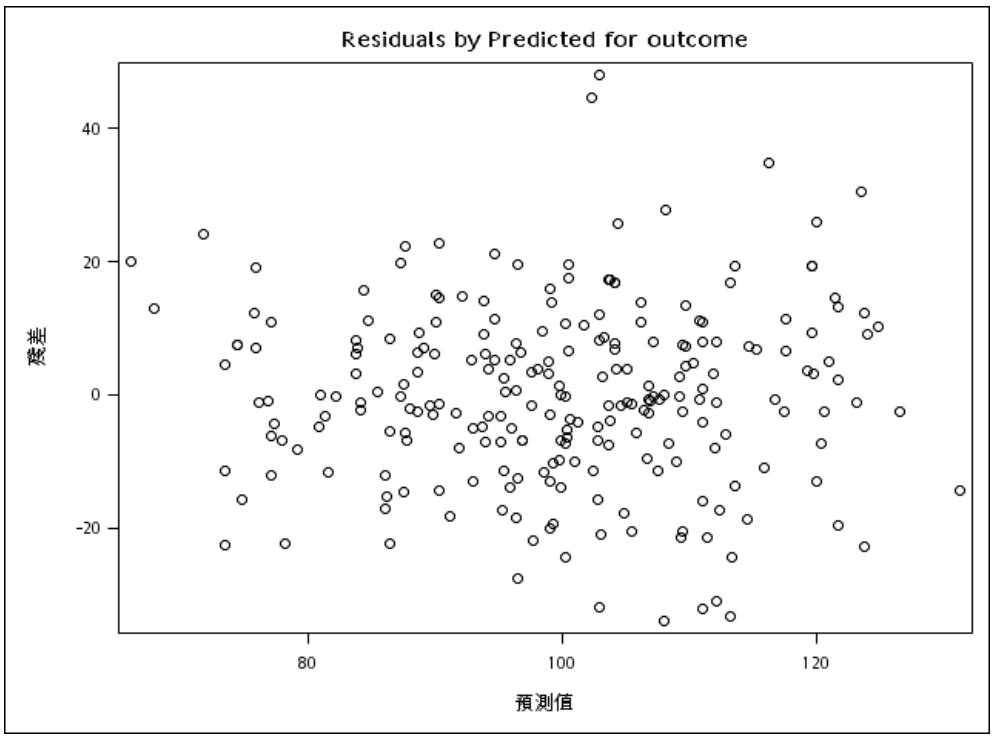


Figure 10: The residual plot of variable, outcome

Missing values affects the unbalanced number of samples in each group

- Experimental study design

Internal validity is one of the factors that affect the experiment. It represents whether the possible sources of error can be effectively controlled in the research experiment. Unbalanced samples and the loss of patients may affect the internal validity and make the possible range of error get bigger. Among them, if the sample is unbalanced, it may cause the trial design to be interfered by other non-research factors. Therefore, in addition to the difference in the experimental factors of the study, other factors between the groups should maintain the same distribution to avoid that the results of the study will not be distorted. Disturbed by non-research factors.

- Analysis

Imbalance in the number of samples can cause misleading results, which in turn can lead to biased results. An unbalanced number of samples can easily lead to false negative or false positive, which means type I error and type II error. In clinical trials, the harm caused by false negatives is sometimes irreversible, so type II error should be avoided. Occurs, so the data imbalance should be resolved.

Conclusion

According to the above statistical methods, different doses of drugs, gender, and the two interactions have no significant effects on the patients.

Appendix

Appendix Table 1: Information of data in Postoperative pain Clinical Trial

Dimensions: 122 x 26
Duplicates: 0

Variable	Stats / Values	Freqs (% of Valid)	Graph	Missing
id [integer]	Mean (sd) : 66.2 (36.9) min < med < max: 1 < 65.5 < 132 IQR (CV) : 62.5 (0.6)	122 distinct values		0 (0%)
treat [character]	1. A 2. B 3. C	41 (33.6%) 41 (33.6%) 40 (32.8%)		0 (0%)
y1 [integer]	Mean (sd) : 3 (1.8) min < med < max: 0 < 3 < 8 IQR (CV) : 2 (0.6)	0 : 8 (6.6%) 1 : 17 (13.9%) 2 : 25 (20.5%) 3 : 28 (22.9%) 4 : 25 (20.5%) 5 : 8 (6.6%) 6 : 5 (4.1%) 7 : 2 (1.6%) 8 : 4 (3.3%)		0 (0%)
y2 [integer]	Mean (sd) : 2.4 (2) min < med < max: 0 < 2 < 8 IQR (CV) : 2 (0.8)	0 : 26 (21.3%) 1 : 19 (15.6%) 2 : 24 (19.7%) 3 : 23 (18.9%) 4 : 12 (9.8%) 5 : 7 (5.7%) 6 : 7 (5.7%) 7 : 3 (2.5%) 8 : 1 (0.8%)		0 (0%)

y3 [integer]	Mean (sd) : 2.1 (1.9) min < med < max: 0 < 2 < 8 IQR (CV) : 2.8 (0.9)	0 : 31 (25.4%) 1 : 29 (23.8%) 2 : 18 (14.8%) 3 : 14 (11.5%) 4 : 16 (13.1%) 5 : 7 (5.7%) 6 : 5 (4.1%) 7 : 1 (0.8%) 8 : 1 (0.8%)		0 (0%)
y4 [integer]	Mean (sd) : 2 (1.8) min < med < max: 0 < 2 < 8 IQR (CV) : 2 (0.9)	0 : 26 (21.3%) 1 : 27 (22.1%) 2 : 31 (25.4%) 3 : 16 (13.1%) 4 : 9 (7.4%) 5 : 8 (6.6%) 6 : 2 (1.6%) 7 : 1 (0.8%) 8 : 2 (1.6%)		0 (0%)
y5 [integer]	Mean (sd) : 2 (1.8) min < med < max: 0 < 2 < 9 IQR (CV) : 2 (0.9)	0 : 27 (22.1%) 1 : 30 (24.6%) 2 : 26 (21.3%) 3 : 12 (9.8%) 4 : 16 (13.1%) 5 : 8 (6.6%) 6 : 1 (0.8%) 7 : 1 (0.8%) 9 : 1 (0.8%)		0 (0%)
y6 [integer]	Mean (sd) : 2 (1.8) min < med < max: 0 < 2 < 7 IQR (CV) : 2 (0.9)	0 : 25 (20.5%) 1 : 32 (26.2%) 2 : 28 (22.9%) 3 : 16 (13.1%) 4 : 7 (5.7%) 5 : 7 (5.7%) 6 : 4 (3.3%) 7 : 3 (2.5%)		0 (0%)
y7 [integer]	Mean (sd) : 1.7 (1.7) min < med < max: 0 < 1 < 7 IQR (CV) : 3 (1)	0 : 38 (31.1%) 1 : 26 (21.3%) 2 : 23 (18.9%) 3 : 16 (13.1%) 4 : 11 (9.0%) 5 : 6 (4.9%) 7 : 2 (1.6%)		0 (0%)

y8 [integer]	Mean (sd) : 1.8 (1.6) min < med < max: 0 < 1 < 7 IQR (CV) : 2.8 (0.9)	0 : 31 (25.4%) 1 : 31 (25.4%) 2 : 23 (18.9%) 3 : 20 (16.4%) 4 : 9 (7.4%) 5 : 6 (4.9%) 6 : 1 (0.8%) 7 : 1 (0.8%)		0 (0%)
y9 [integer]	Mean (sd) : 1.7 (1.6) min < med < max: 0 < 1 < 8 IQR (CV) : 2.8 (1)	0 : 32 (26.2%) 1 : 34 (27.9%) 2 : 25 (20.5%) 3 : 16 (13.1%) 4 : 9 (7.4%) 5 : 4 (3.3%) 8 : 2 (1.6%)		0 (0%)
y10 [integer]	Mean (sd) : 1.8 (1.6) min < med < max: 0 < 1 < 7 IQR (CV) : 2 (0.9)	0 : 24 (19.7%) 1 : 40 (32.8%) 2 : 25 (20.5%) 3 : 17 (13.9%) 4 : 8 (6.6%) 5 : 3 (2.5%) 6 : 4 (3.3%) 7 : 1 (0.8%)		0 (0%)
y11 [integer]	Mean (sd) : 1.7 (1.7) min < med < max: 0 < 1 < 7 IQR (CV) : 3 (1)	0 : 41 (33.6%) 1 : 24 (19.7%) 2 : 22 (18.0%) 3 : 18 (14.8%) 4 : 7 (5.7%) 5 : 6 (4.9%) 6 : 3 (2.5%) 7 : 1 (0.8%)		0 (0%)
y12 [integer]	Mean (sd) : 1.6 (1.5) min < med < max: 0 < 1 < 6 IQR (CV) : 3 (0.9)	0 : 37 (30.3%) 1 : 26 (21.3%) 2 : 25 (20.5%) 3 : 22 (18.0%) 4 : 7 (5.7%) 5 : 4 (3.3%) 6 : 1 (0.8%)		0 (0%)

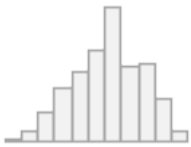
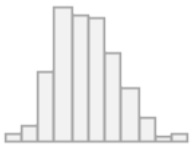
y13 [integer]	Mean (sd) : 1.8 (1.7) min < med < max: 0 < 1.5 < 8 IQR (CV) : 2.5 (0.9)	0 : 31 (25.4%) 1 : 30 (24.6%) 2 : 30 (24.6%) 3 : 15 (12.3%) 4 : 5 (4.1%) 5 : 6 (4.9%) 6 : 3 (2.5%) 7 : 1 (0.8%) 8 : 1 (0.8%)		0 (0%)
y14 [integer]	Mean (sd) : 1.6 (1.5) min < med < max: 0 < 1 < 6 IQR (CV) : 2 (1)	0 : 36 (29.5%) 1 : 30 (24.6%) 2 : 28 (22.9%) 3 : 11 (9.0%) 4 : 10 (8.2%) 5 : 5 (4.1%) 6 : 2 (1.6%)		0 (0%)
y15 [integer]	Mean (sd) : 1.4 (1.3) min < med < max: 0 < 1 < 5 IQR (CV) : 2 (0.9)	0 : 33 (27.1%) 1 : 39 (32.0%) 2 : 26 (21.3%) 3 : 15 (12.3%) 4 : 5 (4.1%) 5 : 4 (3.3%)		0 (0%)
y16 [integer]	Mean (sd) : 1.5 (1.4) min < med < max: 0 < 1 < 6 IQR (CV) : 2 (0.9)	0 : 33 (27.1%) 1 : 35 (28.7%) 2 : 24 (19.7%) 3 : 20 (16.4%) 4 : 7 (5.7%) 5 : 2 (1.6%) 6 : 1 (0.8%)		0 (0%)
y17 [integer]	Mean (sd) : 1.3 (1.4) min < med < max: 0 < 1 < 6 IQR (CV) : 2 (1.1)	0 : 45 (36.9%) 1 : 37 (30.3%) 2 : 20 (16.4%) 3 : 11 (9.0%) 4 : 4 (3.3%) 5 : 3 (2.5%) 6 : 2 (1.6%)		0 (0%)
y18 [integer]	Mean (sd) : 1.1 (1.2) min < med < max: 0 < 1 < 6 IQR (CV) : 2 (1.1)	0 : 48 (39.3%) 1 : 38 (31.1%) 2 : 22 (18.0%) 3 : 11 (9.0%) 5 : 2 (1.6%) 6 : 1 (0.8%)		0 (0%)

y19 [integer]	Mean (sd) : 1.2 (1.3) min < med < max: 0 < 1 < 5 IQR (CV) : 2 (1.1)	0 : 46 (37.7%) 1 : 36 (29.5%) 2 : 21 (17.2%) 3 : 10 (8.2%) 4 : 7 (5.7%) 5 : 2 (1.6%)		0 (0%)
y20 [integer]	Mean (sd) : 1.3 (1.4) min < med < max: 0 < 1 < 7 IQR (CV) : 2 (1.1)	0 : 41 (33.6%) 1 : 38 (31.1%) 2 : 20 (16.4%) 3 : 13 (10.7%) 4 : 6 (4.9%) 5 : 2 (1.6%) 6 : 1 (0.8%) 7 : 1 (0.8%)		0 (0%)
y21 [integer]	Mean (sd) : 1.5 (1.5) min < med < max: 0 < 1 < 8 IQR (CV) : 2 (1)	0 : 36 (29.5%) 1 : 31 (25.4%) 2 : 30 (24.6%) 3 : 13 (10.7%) 4 : 9 (7.4%) 5 : 1 (0.8%) 7 : 1 (0.8%) 8 : 1 (0.8%)		0 (0%)
y22 [integer]	Mean (sd) : 1.5 (1.7) min < med < max: 0 < 1 < 9 IQR (CV) : 2 (1.1)	0 : 45 (36.9%) 1 : 30 (24.6%) 2 : 18 (14.8%) 3 : 13 (10.7%) 4 : 10 (8.2%) 5 : 4 (3.3%) 6 : 1 (0.8%) 9 : 1 (0.8%)		0 (0%)
y23 [integer]	Mean (sd) : 1.5 (1.6) min < med < max: 0 < 1 < 8 IQR (CV) : 2 (1.1)	0 : 40 (32.8%) 1 : 34 (27.9%) 2 : 24 (19.7%) 3 : 14 (11.5%) 4 : 4 (3.3%) 5 : 1 (0.8%) 6 : 3 (2.5%) 7 : 1 (0.8%) 8 : 1 (0.8%)		0 (0%)

y24 [integer]	Mean (sd) : 1.9 (1.8) min < med < max: 0 < 1 < 7 IQR (CV) : 3 (1)	0 : 33 (27.1%) 1 : 32 (26.2%) 2 : 17 (13.9%) 3 : 18 (14.8%) 4 : 9 (7.4%) 5 : 7 (5.7%) 6 : 5 (4.1%) 7 : 1 (0.8%)		0 (0%)

Appendix Table 2: Information of data in CRD Clinical Trial

Dimensions: 240 x 5
Duplicates: 0

Variable	Stats / Values	Freqs (% of Valid)	Graph	Missing
dose [integer]	Mean (sd) : 60 (44.8) min < med < max: 0 < 60 < 120 IQR (CV) : 60 (0.7)	0 : 60 (25.0%) 40 : 60 (25.0%) 80 : 60 (25.0%) 120 : 60 (25.0%)		0 (0%)
id [integer]	Mean (sd) : 120.5 (69.4) min < med < max: 1 < 120.5 < 240 IQR (CV) : 119.5 (0.6)	240 distinct values (Integer sequence)		0 (0%)
gender [character]	1. Female 2. Male	124 (51.7%) 116 (48.3%)		0 (0%)
baseline [integer]	Mean (sd) : 86.3 (10.3) min < med < max: 58 < 87 < 108 IQR (CV) : 15 (0.1)	47 distinct values		0 (0%)
outcome [integer]	Mean (sd) : 99.7 (18.7) min < med < max: 51 < 98.5 < 154 IQR (CV) : 25 (0.2)	71 distinct values		0 (0%)